Course guide
270960 - OTDM - Optimization Techniques for Data Mining

Last modified: 14/07/2025

Unit in charge: Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research.

Degree: MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).

Academic year: 2025 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer: JORDI CASTRO PÉREZ

Others: Primer quadrimestre:
JORDI CASTRO PÉREZ - 10
FRANCISCO JAVIER HEREDIA CERVERA - 10
JESSICA RODRÍGUEZ PEREIRA - 10

PRIOR SKILLS

Basic background in linear algebra, calculus, and programming languages is needed for the course.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science

Generical:
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats

Transversal:
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Basic:
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

TEACHING METHODOLOGY

The students will have available all the course material.

About two thirds of lecture time will be devoted to optimization algorithms and their properties, and the rest will be for presenting and solving exercises and problems

Lab sessions will be devoted to the solution of some data science applications (neural networks, support vector machines, clustering) using optimization methods.

LEARNING OBJECTIVES OF THE SUBJECT

1. Discerning what is an optimization problem and its type and having a basic knowledge of optimization algorithms
2.Optimization of neural networks and support vector machines.
3. Modelling languages. Formulating optimization problems and representing them through a modeling language. Choosing and adequate solver for a given problem.

STUDY LOAD

Type	Hours	Percentage
Self study	96,0	64.00
Hours large group	54,0	36.00

Total learning time: 150 h

CONTENTS

Unconstrained Optimization

Description:
Optimality conditions. Convexity. Descent directions.
Line search. Acceptability of step sizes.
General minimization algorithm.
Gradient method. Rate of convergence.
Newton's method. Factorizations to ensure convergence.
Quasi-Newton methods.
Optimization of neural networks.

Constrained Optimization and Support Vector Machines.

Description:
- Introduction to the modelling langauge AMPL.
- Introduction to Support Vector Macines (SVM)
- KKT Optimality conditions of constrained optimization. Optimality conditions of SVM.
- Duality in Optimization. The dual of the SVM.

Integer Programming

Description:
- Modelling problems with binary variables.
- The branch and bound algorithm for integer programming
- Gomory's cutting planes algorithm.
- Minimal spanning tree problem and algorithms of Kruskal and Prim. Application to clustering.

ACTIVITIES

Unconstrained Optimization

Description:
Optimality conditions. Convexity. Descent directions.
Line search. Acceptability of step sizes.
General minimization algorithm.
Gradient method. Rate of convergence.
Newton's method. Factorizations to ensure convergence.
Weighted least squares.
Introduction to AMPL. The Neos solver site.

Specific objectives:
1, 2, 3

Related competencies :
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 50h 18m
Self study: 33h
Theory classes: 17h 18m

Constrained Optimization and Support Vector Machines

Description:
- Introduction to Support Vector Macines (SVM)
- KKT Optimality conditions of constrained optimization. Optimality conditions of SVM.
- Duality in Optimization. The dual of the SVM.

Specific objectives:
1, 2, 3

Related competencies :
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 50h 24m
Self study: 33h
Theory classes: 17h 24m

Integer Programming

Description:
- Modelling problems with binary variables.
- The branch and bound algorithm for integer programming
- Gomory's cutting planes algorithm.
- Minimal spanning tree problem and algorithms of Kruskal and Prim

Specific objectives:
1, 3

Related competencies :
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CG2. Identify and apply methods of data analysis, knowledge extraction and visualization for data collected in disparate formats
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.

Full-or-part-time: 50h 18m
Self study: 33h
Theory classes: 17h 18m

GRADING SYSTEM

Each of the three parts of the topic (unconstrained optimization, constrained optimization, integer optimization) has a weight 1/3 on the final mark, and is evaluated as: 40% theory and 60% practical assignment. The evaluation activities are:

- Theory (40%). There will be two midterm exams (there is no final exam). The first midterm exam (2/3 of 40%) will consist on some practical exercises about the first two parts of the course, and it will be done before January. The second midterm exam (1/3 of 40%) will be about the third part of the course, and it will be done in January.

- Practical assignments (60%). There will be 3 lab assignments, one for each part of the course, all of them with the same weight on the final mark.

Additional exercises provided during the lectures may be taken into consideration to decide or to boost the final mark.

BIBLIOGRAPHY

Basic:
- Nocedal, J.; Wright, S.J. Numerical optimization. 2nd ed. Berlin: Springer, 2006. ISBN 9780387303031.
- Luenberger, D.G.; Ye, Y. Linear and nonlinear programming. 5th ed. Cham: Springer, 2021. ISBN 9783030854492.
- Wolsey, L.A. Integer programming. 2nd ed. Hoboken, New Jersey: Wiley, 2021. ISBN 9781119606536.
- Fourer, R.; Gay, D.M.; Kernighan, B.W. AMPL: a modeling language for mathematical programming. 2nd ed. Belmont, CA: Thomson Brooks/Cole, 2003. ISBN 0534388094.
- Cristianini, N.; Shawe-Taylor, J. An introduction to support vector machines: and other kermel-based learning methods. New York: Cambridge University Press, 2000. ISBN 0521780195.

RESOURCES

Hyperlink:
- http://www-eio.upc.es/teaching/ple/pfc_ing.html- https://neos-server.org/neos/